

The Tools for Machine Learning

Many tools have been developed to make machine learning more powerful and easier to implement. On this page, we'll take a look at the typical components you might employ in a machine learning ecosystem. You don't need to understand the details of these tools at this stage, and we don't assume you've had previous experience with them. Our goal at this point is simply to give you some idea of what some of the popular tools are and how they relate to one another.



The Machine Learning Ecosystem

A typical machine learning ecosystem is made up of three main components:

- 1. Libraries.** When you're working on a machine learning project, you likely will not want to write all of the necessary code yourself—instead, you'll want to make use of code that has already been created and refined. That's where libraries come in. *A library* is a collection of pre-written (and compiled) code that you can make use of in your own project. *NumPy* is an example of a library popularly used in data science, while *TensorFlow* is a library specifically designed for machine learning. Read this [article](#) for some other useful library.
- 2. Development environments.** A *development environment* is a software application (or sometimes a group of applications) that provide a whole suite of tools designed to help you (as the developer or machine learning engineer) build out your projects. *Jupyter Notebooks* and *Visual Studio* are examples of development environments that are popular for coding many different types of projects, including machine learning projects.

3. Cloud services. A *cloud service* is a service that offers data storage or computing power over the Internet. In the context of machine learning, you can use a cloud service to access a server that is likely far more powerful than your own machine, or that comes equipped with machine learning models that are ready for you to use. Read more information about different cloud services from this [article](#)

For each of these components, there are multiple options you can choose from. Let's have a look at some examples.

Notebooks

Notebooks are originally created as a documenting tool that others can use to reproduce experiments. Notebooks typically contain a combination of runnable code, output, formatted text, and visualizations. One of the most popular open-source notebooks used today by data scientists and data science engineers is [Jupyter notebook](#), which can combine code, formatted text (markdown) and visualization.

Notebooks contains several independent **cells** that allow for the execution of code snippets within those cells. The output of each cell can be saved in the notebook and viewed by others.

End-to-end with Azure

You can analyze and train a small amount of data with your local machine using Jupyter notebook, Visual studio, or other tools. But with very large amounts of data, or you need a faster processor, it's a better idea to train and test the model *remotely* using *cloud services* such as Microsoft Azure. You can use Azure Data Science Virtual Machine, Azure Databricks, Azure Machine Learning Compute, or SQL server ML services to train and test models and use Azure Kubernetes to deploy models.

QUIZ QUESTION

Below are the development environments we just discussed. Can you match each one with its description?

Submit to check your answer choices!

DESCRIPTION	DEVELOPMENT ENVIRONMENT
Microsoft's core development environment	Visual Studio
Open-source tool that can combine code, markdown, and visualizations together in a single document.	Jupyter Notebooks
A light-weight code editor from Microsoft	Visual Studio Code
Data analytics platform, optimized for use with Microsoft cloud services	Azure Databricks

SUBMIT